
Rule DAS600: Excessive Control Area (CA) splits occurred

Finding: CPExpert determined that an excessive number of Control Area (CA) splits occurred for the VSAM data sets listed.

Impact: This finding is used to assess problems or potential problems with VSAM Control Area splits for the data sets listed.

Discussion: A VSAM file structure consists of one or more *Control Intervals (CIs)* and one or more *Control Areas (CAs)*.

- A **Control Interval** is a continuous area of direct access storage that VSAM uses to store logical records. When a logical record is read from direct access storage, the entire Control Interval containing the record is read into a VSAM buffer in virtual storage. The desired logical record is then transferred from the VSAM buffer to a user-defined buffer or work area. While logical records within a Control Interval may vary in length, all Control Intervals in a specific VSAM data set are of the same length.

In addition to the logical records, a Control Interval consists of free space, and control information. The free space initially is unused, and is used to accommodate inserted logical records or for changes in the length of logical records. The control information describes the amount and location of free space, and describes the length of records and how many adjacent records are of the same length.

- A **Control Area** contains one or more Control Intervals. The Control Intervals are grouped together into fixed-length contiguous areas of direct access storage. A VSAM data set consists of one or more Control Areas.

The minimum size of a Control Area is one track of DASD storage, and the maximum size of a Control Area is one cylinder. The size of a Control Area is not specified by a user; the size of the CA is *calculated* by VSAM based on the amount of space allocated to a VSAM data set.

Logical records can be inserted into a VSAM keyed sequenced data set (KSDS) or a variable relative record data set (VRRDS), or a record length can be increased. When either of the actions occur, space must be available in a Control Interval to accommodate the new record (or accommodate the increased length). The space for the new record (or

increased length) normally is obtained from the *free space* that was allocated to the Control Interval¹ when the data set was loaded.

If enough free space is not available in the Control Interval, a *Control Interval Split* occurs. When a Control Interval is split, approximately one-half of the logical records in the original Control Interval will be transferred to a new Control Interval, and the records are then deleted in the old Control Interval. Space for the new Control Interval is obtained from the free space that is associated with the Control Area. Thus, inserted records or longer records are inserted into Control Intervals using the free space associated with the Control Interval. If the Control Interval does not have sufficient space, a new Control Interval will be created, using free space in the Control Area² to which the original Control Interval belongs.

If enough free space is not available in the Control Area, a *Control Area Split* occurs. When a Control Area is split, approximately one-half of the Control Intervals from the Control Area are moved to the end of the data set. The records are then deleted from the old Control Area (requiring additional I/O operations). For example, if the CA is one cylinder, then approximately one-half of the records in the cylinder is moved to the end of the data set, and the records that were moved are then deleted from their location in the original cylinder. Depending on the size of the Control Area, considerable overhead³ can result from Control Area splits.

Additionally, locking may be involved during Control Area split processing, depending on the SHARE options associated with the VSAM data set. The below table shows the level at which locking occurs, and the condition that causes the lock level.

LOCK LEVEL	CONDITION
Control Interval	Adding a record or updating a record in place, without causing a Control Interval split
Control Area	Adding a record or updating a record in place, causing a Control Interval split, but not causing a Control Area split.
Data Set	Adding a record or updating a record in place, causing a Control Interval split, and causing a Control Area split.

¹The amount of free space reserved for each Control Interval is controlled by the FREESPACE parameter. The first value of the FREESPACE keyword specifies a percent of each Control Interval that is to be reserved for free space.

²The amount of free space reserved for each Control Area is controlled by the FREESPACE parameter. The second value of the FREESPACE parameter specifies a percent of each Control Area that is to be reserved for free space.

³Note that the overhead occurs only during the CA split. There is essentially no additional overhead after the CA split.

Considering the overhead involved and the locking that can occur, Control Area splits should be avoided if possible.

CPEXpert examines the SMF Type 64 information contained in MXG TYPE64 data set to identify VSAM data sets that have excessive Control Area splits.

CPEXpert sums the ACCASPLT variable (the number of CA splits since the data set was created) and the CASPLITS variable (the number of CA splits with the current OPEN of the data set). CPEXpert compares this sum with the **CASPLITS** guidance variable in USOURCE(DASGUIDE). CPEXpert produces Rule DAS600 when the total number of CA splits exceeds the value specified by the **CASPLITS** guidance variable.

The default value for the **CASPLITS** guidance variable is 10, indicating that CPEXpert should produce Rule DAS600 when a VSAM data set experienced more than 10 CA splits.

The following example illustrates the output from Rule DAS600:

RULE DAS600: EXCESSIVE CONTROL AREA SPLITS OCCURRED

VOLSER: RLS014. More than 10 Control Area (CA) splits occurred for the VSAM data sets listed below. CA splits cause considerable overhead during the split, and CA splits should be avoided.

SMF TIME STAMP	JOB NAME	VSAM DATA SET	TOTAL CA SPLITS	CA SPLITS THIS OPEN
10:30,29AUG2001	CICS2AGC	RLSADSW.VF03D.DATAENDB.DATA.....	115	9
10:30,29AUG2001	CICS2AGC	RLSADSW.VF04D.DATAENDB.DATA.....	117	11
10:30,29AUG2001	CICS2AGA	RLSADSW.VF07D.ITEMACT.DATA.....	63	49

Although not shown in this example, CPEXpert also reports the total number of inserts to the VSAM data set and inserts for the current OPEN of the VSAM data set.

Suggestion: The action taken to reduce CA splits requires a reorganization of the data set, often with different values for the FREESPACE parameter. If CA splits occur frequently for a VSAM data set, you should consider the following alternatives:

- **Reduce CI free space and increase CA free space.** Unless inserts (or increases in record length) are evenly distributed across Control Intervals, you should reduce (or eliminate) the Control Interval FREESPACE amount and increase the Control Area FREESPACE amount. Control Interval splits do not cause much overhead (little data is moved and the movement remains within the Control Area). Reducing

or eliminating the Control Interval FREESPACE amount eliminates wasted DASD space, unless inserts take place for most Control Intervals. Adding free space to the Control Area will often significantly reduce the number of Control Area splits and significantly reduce the overhead associated with the Control Area splits.

If new records will be evenly distributed throughout the data set, control area free space should equal the percentage of records to be added to the data set after the data set is loaded (specify FREESPACE (0 nn), where nn equals the percentage of records to be added.)

Unfortunately, simply increasing the amount of Control Area free space might not be a good solution. Specifying too much free space can result in more direct access storage required to contain the data set and much of this space might be wasted if inserts are clustered.

- **Allow the insert pattern to control where splits occur.** If insertions will be unevenly distributed throughout the data set and you cannot tell where the insertions are likely to occur, you can specify a small amount of free space. In this case, splits will occur *in those areas where insertions occur*. This option reduces the amount of wasted DASD space caused by a large amount of unused free space.
- **Allocate CA free space where clustering occurs.** If you can determine the clustering nature of the record insertions and the data set is large, you might consider loading the data set in stages⁴, with free space specified in areas where clustering is expected.
- **Spread the overhead.** If the VSAM data set is used **primarily by on-line applications**, you might specify a relatively small primary allocation, so that the impact of Control Area splits would be spread across the transactions encountering the Control Area splits. If a Control Area were one cylinder, the considerable overhead of splitting the Control Area would be incurred by the transaction that caused a Control Area split to occur. This might be a relatively random event (depending on the characteristics of other transactions and which user submitted the transactions), and would result in unpredictably poor response time. One way to reduce the impact on any particular on-line user would be to make the Control Area relatively small (for example, use a *primary allocation unit* of only a track, which would cause the Control Area to be only a track) so that when a CA split did occur, the immediate impact would not be as significant.

⁴ IBM shows an example of this approach in *DFSMS: Using Data Sets* (Section 2.5.3.2: Altering the Free Space Specification When Loading a Data Set)

-
- **You can not determine the clustering nature of the record insertions and the data set is relatively small.** If this is the situation, you probably should ignore the CA splits reported by DAS600, since the CA splits should not continue to occur. If the CA splits **should** continue to occur, you probably will see the data set allocation begin to acquire additional extents. This is because the size of the CA is determined by VSAM based on the primary and secondary allocation. In this situation, CPExpert will produce Rule DAS604 (Excessive secondary extents were allocated).
 - If none of the above actions are appropriate, you can change the **CASPLITS** guidance variable in USOURCE(DASGUIDE). Section 3 describes how to change the CASPLITS guidance variable if you feel that Rule DAS600 is produced too often, or if you do not wish to take action when only 10 CA splits occur.
 - Alternatively, you can exclude the reported VSAM data sets from analysis. Section 3 describes how to exclude VSAM data sets from analysis. However, you should be aware that no analysis of potential VSAM problems will be performed on data sets that are excluded from analysis.

Reference: *DFSMS: Using Data Sets* (SC26-7339 for OS/390; SC26-7410 for z/OS)
Section 2.11.4.2.3: Control Interval Splits
Section 2.5.2: Optimizing Control Area Size
Section 2.5.3.1: Selecting the Optimal Percentage of Free Space
Section 2.5.3.2: Altering the Free Space Specification When Loading a Data Set